图书情报工作
第63卷 第3期 2019年2月

数据驱动下数字图书馆用户画像模型构建*

许鹏程[1] 毕强[1] 张晗[1] 牟冬梅[2]

[1] 吉林大学管理学院 长春 130022 [2] 吉林大学公共卫生学院 长春 130021

摘要：［目的/意义］为了挖掘用户数据背后隐藏的价值，全面了解用户需求，构建用户画像模型，为数字图书馆实现精准服务提供新动能。［方法/过程］针对数字图书馆用户画像的内涵及特征进行剖析，分析用户画像的数据来源及采集处理过程，提出数据驱动下用户画像数据化→标签化→关联化→可视化的驱动主路线，从自然维度、兴趣维度、社交维度，构建多维度、多层级、立体化的用户画像模型。［结果/结论］详细阐述数字图书馆用户画像模型的构建流程，设计用户画像的框架模型，并将用户画像应用于数字图书馆的精准推荐、个性化检索、精准宣传以及参考决策中，以促进数字图书馆的知识服务升级。

关键词：数据驱动 数字图书馆 用户画像

分类号：G251

DOI：10.13266/j.issn.0252-3116.2019.03.004

国内外学者用新一代信息技术将当今时代直接概括和描述为“大数据时代”“智能时代”“算法时代”，图书馆所依赖的知识创造与利用环境正在从信息时代进入到“数据时代”，大数据驱动的“数据化”浪潮使图书馆拥有多种形态的资源和多元化的数据，多种形态资源和多样化的数据构成了可充分集成关联的数字图书馆服务的大数据环境[1]。面对图书馆拥有的海量的数字资源，用户却无法快速而精准地找到自己感兴趣的资源，使用户陷入知识迷航的困境中。鉴于此，数字图书馆亟需从信息服务向知识服务转型，而构建动态用户画像是深刻理解用户需求、实时洞察用户偏好、实现服务转型、解决用户精确需求与数字图书馆粗放型服务不对称问题的有效途径。

1 相关研究概述

1.1 用户画像研究概述

交互设计之父 A. Cooper 最早提出了用户画像（persona）这一概念，指出用户画像是建立在一系列真实数据之上的用户目标模型，是对真实用户的虚拟化[2]。用户画像最早被应用在交互设计或产品设计领域中，用户画像是指对特定服务群体真实特征的勾勒，是刻画目标用户、联系用户诉求等的一种有效工具[3]。早期用户画像模型一般是基于虚拟视角构建的，设计师将头脑中的主观想象具体化为目标客户的轮廓，从而设计出产品原型，但基于虚构视角构建的用户画像完全取决于设计师的主观假设，容易产生误导[4]。随着互联网技术的发展和大数据时代的来临，在数据驱动下，用户画像的内涵和外延都发生了变化，主要是通过数据刻画用户特征，从而为用户提供优质服务。因此用户画像有了另一种表达：user profile，本文即采用该表达。同时，构建用户画像的视角由虚构视角向目标导向的视角、角色视角、参与视角转变，更加注重采集用户数据以支撑画像结果。C. Teixeira 等认为用户画像是从海量数据汇中通过提取出个人信息集合，来独立进行用户需求、偏好和兴趣描述的模型[5]。

用户画像的构建离不开算法与技术的支持，正如 L. Zeng 等所言，运用大数据技术和数据挖掘算法构建用户画像模型的研究较多，如向量空间模型、主题模型、神经网络模型等算法[6]。A. Sandra 等[7]提出了一种基于上下文偏好规则的概念构建用户画像的方法，以更准确地反映用户偏好特征。A. Farida 等[8]提出了一种基于动态贝叶斯网络方法从用户与搜索系统的交

* 本文系国家自然科学基金面上项目“嵌入式知识服务驱动下的领域多维知识库构建”（项目编号：71573102）研究成果之一。

作者简介：许鹏程（ORCID：0000-0001-5519-8550），硕士研究生；毕强（ORCID：0000-0001-7381-4986），教授，博士生导师；张晗（ORCID：0000-0002-4586-0819），硕士研究生；牟冬梅（ORCID：0000-0003-0237-034X），教授，博士生导师，通讯作者，E-mail：moudm@jlu.edu.cn。

收稿日期：2018-07-06 修回日期：2018-10-10 本文起止页码：30-37 本文责任编辑：王传清

互中构建用户画像。J. Yu 等[9]提出了一种基于语义和浏览顺序构建和更新用户画像的方法,并引入了认知心理学的记忆模型,以保证用户画像的动态性。由此可见,用户画像已经整合和利用了许多成熟的算法和技术，但究其本质,其实是多维标签组合的建模[10]，通过从用户多源数据中提取用户标签的形式来进行用户画像,是构建用户画像模型的核心。

随着大数据和"互联网＋"理念的深入,企业的关注点日益聚焦于怎样利用大数据来实现精准服务,通过用户画像可以帮助产品经理精准地了解和预测用户需求,从而精准定位客户群体,进行宣传和个性化推荐,最终实现精准服务,因此,用户画像在各领域得到广泛应用,尤其是电子商务领域[11],用户画像成为实现精准营销和决策的依据。另外,在旅游[12]、金融[13]、新闻[14]、社交网络[15]、健康[16]等领域,用户画像同样发挥重要作用,甚至已成为众多领域实现精准服务、精准营销的突破点。

1.2 图书馆领域用户画像研究现状

在图书馆领域用户画像研究已成为研究热点,用户画像在国外图书馆中的应用最早出现在 1985 年,英国国家书目和 Blaise-line 通过电话采访和个人访谈的方式调查用户对英国国家书目和 Blaise-line 的使用情况,并形成相关分析,以此来进行服务优化[17]。G. Amato 等[18]详细阐述了用户画像在数字图书馆领域的应用,认为信息提供者的最终目标是满足用户信息需求,开展个性化服务依赖表示用户信息需求的用户画像,因此数字图书馆需要构建用户画像模型。G. Amato 等[19]还就数字图书馆个性化服务中用户画像的结构和机制展开研究,详细介绍了用户画像的表示、构建和更新等关键问题。随着用户画像研究的深入,国外众多学者将用户画像主要应用于图书馆的信息推荐[20]、信息过滤[21]、服务设计[22]等领域中。

国内图书馆领域对用户画像的研究起步较晚,近 3 年才逐渐成为研究热点,研究成果也相对较少,主要集中在图书馆用户画像的模型构建和实践应用。刘速[23]、杨帆[24]从具体实践案例中阐述用户画像的构建思路,并提出了具体的用户画像分析方法。何娟[25]基于用户画像构建图书馆智慧推荐系统,以提高阅读推广效率,实现精准推荐服务。尹相权、薛欢雪等探索了高校图书馆用户画像模型并进行了实证研究,分析高校学生使用行为[26],优化高校图书馆学科服务[27]。另外还有部分学者将用户画像应用于数字图书馆的知识社区[28]和问答社区[29]中,从用户微观数据进行用户画像标签建模;范晓玉等[30]则融合多源数据从基础属性、科研偏好和科研关系 3 个维度提取科研人员信息标签,并以可视化的方式展示科研人员画像,这些研究将用户画像应用范围得更加细化、更加深入,颇具实践意义。

综上所述,国外图书馆领域关于用户画像的研究起步较早,从理论基础,到模型构建,再到方法与技术,最后到实践应用,其研究较为成熟,而国内图书馆领域关于用户画像的研究还有待丰富和完善,因此可以借鉴国外成熟的用户画像模型,并结合我国图书馆发展现状和用户群体特征构建完备的图书馆用户画像模型,加强对用户画像的实践和应用,从而更好地把握用户需求，真正实现精准服务。另外,在图书馆用户画像的研究中,对用户隐私的研究较少,毕竟用户画像的构建需要大量的用户数据,隐私问题不可忽视,未来的研究中应将用户隐私问题纳入重要考虑因素。

2 数字图书馆用户画像的内涵及特征

2.1 数字图书馆用户画像的内涵

数字图书馆用户画像是图书馆为了深入了解用户特征、预测用户真实需求、激发用户潜在需求等,在一系列真实数据的基础之上通过描述用户特征、需求和偏好,构建的目标用户模型,是用户信息面貌的虚拟刻画,其目标是实现精准服务。其中,真实数据是指能多角度的反映用户特征的数据,如用户背景、用户兴趣、用户习惯、用户行为等[31]。数字图书馆应基于参与视角,采用数据－用户标签映射法构建用户画像。数据－用户标签映射法是典型的数据驱动的用户画像,数字图书馆用户画像的过程可表现为数据化→标签化→关联化→可视化,即首先要采集用户相关数据,对其进行预处理,实现数据化;基于用户的基本属性和行为数据将用户画像标签化,建立用户标签体系;基于用户互动数据,建立用户间的联系,建立用户关系图谱,实现关联化;最后通过可视化工具将用户画像可视化输出,完成最终的用户画像,并将用户画像数据应用于个性化检索、精准推荐、用户聚类与精准宣传、图书馆参考决策等。

2.2 数字图书馆用户画像的特性

D. Travis 认为用户画像应具有 7 个特性:基本性(primary research)、移情性(empathty)、真实性(realistic)、独特性(singular)、目标性(obejectives)、数量(number)、应用性(applicable)[32]这 7 个特性是基于早期用户画像的内涵提出的,大数据时代背景下用户画

像应更具时代特色,笔者认为数据驱动下数字图书馆用户画像应具有可迭代性、时效性、区隔性、交互性、知识性和聚类性。

2.2.1 可迭代性　用户画像是用户相关数据标签化、关联化、可视化的结果,用户数据由静态数据和动态数据两部分组成。数字图书馆用户静态信息(如姓名、年龄、读者ID等)主要涉及用户基本属性特征,相对稳定;动态信息(如点击、阅读、下载等)主要是通过用户与数字图书馆交互产生,会随着时间的推移而持续累加,用户与系统的每次交互都可能使用户画像结果发生变化,表明用户画像是动态变化的。数字图书馆用户画像应具有可迭代性,能针对用户的需求和行为变化更新用户画像,并相应地调整服务方式。

2.2.2 时效性　受学习计划、认知加深、任务调整、环境变化、信息过载以及时间推移等众多因素的影响,会发生用户兴趣漂移的现象。即使针对同一内容,用户对其感兴趣的程度也可能逐渐提高或者降低甚至消失,用户兴趣漂移现象决定了用户画像要具有时效性。数字图书馆用户画像是某个时间段内目标用户的立体刻画,其准确性在对应的时间段内有效。因此,精准动态的用户画像模型应能够实时准确地追踪用户的兴趣漂移,及时地针对用户兴趣变化做出反应,实时更新用户画像结果。如果用户画像结果有延迟,则可能导致用户画像结果准确性降低。

2.2.3 区隔性　用户画像是具有显著特征的用户模型,是真实用户某个层面、某些维度的特征数据化重组后的虚拟体现。用户画像刻画的用户特征,不是全体用户的平均化特征,而是具有区隔性和对象针对性。刻画出来的用户特征之间一般不会出现重叠和交叉现象。所有特征都是用户画像中独具特色的一员,各种特征各有所专、各有侧重,而又互补共存,形成了用户画像这个汇聚了不同特征维度的集合。用户画像的区隔性使得数字图书馆能够精准地识别出不同用户的行为偏好和动机,为其进一步提升服务指明方向。

2.2.4 交互性　数字图书馆与用户交互的数据是用户画像的重要构成要素,用户交互行为主要体现在用户基于不同的目的和动机使用数字图书馆,由此产生兴趣型交互、问题型交互、社交型交互和利益型交互等,从而利用各种移动媒体,通过操作层面或语义层面与图书馆系统、咨询馆员、用户间进行有传送有反馈的行为互动过程。用户与数字图书馆不断地交互过程中,持续产生大量的用户数据,丰富的用户数据是用户画像可行性和准确性的基础。用户画像是挖掘大量用户数据并将用户信息面貌可视化的结果,其结果可能与用户的真实面貌有所不同,因此数字图书馆在构建用户画像系统过程中需建立完善的用户反馈机制,允许用户对用户画像的结果提出反馈意见,数字图书馆及时整理用户反馈和建议,完善和改进用户画像结果。

2.2.5 知识性　用户既是数据资源的获取者,也是数据的创造者,而数据作为一种资源,具有潜在的知识待被挖掘、被发现和被利用,数字图书馆用户画像的目的在于对用户数据的挖掘及知识创造,从而满足用户碎片化、精细化、个性化的知识需求,因此数字图书馆用户画像应具备强大的应用功能,辅助于数字图书馆的知识发现服务。数字图书馆用户画像是建立在资源画像的基础上,数据驱动下数字图书馆对资源进行细粒度分割,实现数据的碎片化,通过语义关联,最终形成可视化资源画像;数字图书馆从用户角度出发,注重用户参与交互,将用户需求和行为数据同资源关联起来,从而使得用户画像更具应用性和知识性。另外,数字图书馆有一些用户也是知识的创造者,是各领域的专家学者,数字图书馆用户画像应具备建立专家型用户、学术型用户等个人知识库的能力,并建立用户与用户间的联系,促进知识传播与分享,从而实现知识再造,发挥知识价值。

2.2.6 聚类性　用户与用户之间虽然具有区隔性,但也存在共性,正是由于共性的存在,便于聚类同质化的群体、区分异质化对象。用户的相关数据背后隐藏了用户的共性特征,数字图书馆用户画像基于对用户相关数据的分析挖掘,根据用户使用习惯、兴趣偏好、活跃程度、参与度以及影响力等划分标准,实现对用户的分类、聚类,形成不同的用户群体,实现对用户进行分级管理,使得对用户的管理更加科学高效,并针对不同的用户群体进行个性化服务、精准宣传等。

3 数字图书馆用户画像模型构建

3.1 数据采集及处理

用户画像是为了全方位立体化的刻画用户,因此需要以广泛的图书馆用户数据为基础进行画像,多源用户数据是用户画像刻画准确性的保障[33]。T. P. Guimaraes[34]将用户画像的数据来源归纳为:用户的基本素养、学历层次、社会关系、工作状况、位置情况、时间信息等。T. Laouge、J. P. LARDY 和 N. B. ABDALLAH[35]认为信息检索系统中的用户特征信息主要包含两方面:一是与用户相关的稳定因素,如用户的个人信息和行为习惯;二是检索环境下的可变信息,如

搜索目标。围绕上述两种特征信息,构建用户画像模型的数据来源包括用户状态、用户背景、用户目标、用户对相关领域的认知、用户对系统的熟悉程度 5 类。笔者认为数字图书馆用户画像的数据主要包括用户人口统计学数据、用户行为数据、用户社交数据、用户其它数据。其中用户人口统计学数据主要包括读者 ID、姓名、性别、年龄、学历、专业、职业、职称、城市以及邮箱地址等;用户行为数据主要包括登录次数、浏览时长、页面滚动、点击、跳转、下载、收藏、复制,手机端的手势滑动、拖动、放大缩小[36],借阅、检索、咨询等;用户社交数据包括点赞、分享、评论、讨论、互动、关注、引用、被引、合作等;用户其他数据有手机和电脑型号、使用的操作系统、客户端版本、网络类型以及学术成果等。

数字图书馆的用户数据通常由结构化数据、半结构化数据和非结构化数据组成,结构化数据相对比较好采集,并且便于形成用户标签,如用户基本属性数据,可以通过用户注册信息采集。而半结构化数据和非结构化数据量庞大,同时是用户画像的主要数据,如页面浏览、点击、下载等用户行为数据主要存储于用户 Web 日志中,需要通过网络爬虫和日志挖掘技术进行提取。另外,数字图书馆系统中需要嵌入用户页面行为监控插件,在不影响用户的正常使用下收集用户行为数据。用户社交数据是通过用户间的互动、合作等方式,建立起用户 - 用户间、用户 - 专家间的联系来收集。用户相关数据的采集需要借助网页爬虫、Web 日志挖掘等工具,并将采集后的数据进行清洗、转换、规约、集成等预处理,形成有效的用户画像数据。

3.2 数字图书馆用户画像的维度

张慧敏和辛向阳[37]从交互设计的角度提出了构建用户画像的 4 个维度,分别为自然条件维度、价值取向维度、行为习惯维度和认知特征维度,每个维度又包含多个子维度,具有较好的概括性和全面性;陈志明和胡震云[38]构建了用户基本属性、社交属性、兴趣属性和能力属性 4 个维度的 UGC 网站用户画像模型;王凌霄、沈卓和李艳[29]从用户资历、用户参与度、用户回答质量以及用户发展趋势 4 个方面构建了社会问答社区的用户画像,胡媛和毛宁[28]从读者基本信息、用户兴趣爱好、用户活跃度 3 个标签维度建立了数字图书馆知识社区用户画像模型,以上研究对本文的研究颇具指导和借鉴意义。根据数字图书馆的用户属性和需求,笔者认为数字图书馆的用户画像应从自然维度、兴趣维度、社交维度 3 个方面刻画,构建三维多级标签体系的数字图书馆用户画像模型,其中自然维度是基础、兴趣维度是主体、社交维度是价值导向,三者的内在关系是:自然维度是兴趣维度和社交维度的基础,用户的基本属性是催生用户兴趣的形成与变化以及提供社交关系发展的基础;兴趣维度是用户画像的需求,是用户个人偏好和需求的体现,社交维度是用户自然维度和兴趣维度共同作用的结果,同时社交维度也会反作用于兴趣维度,影响用户的兴趣偏好。

3.2.1 自然维度　自然维度旨在对用户进行最基本的了解和刻画,主要基于用户的人口统计学数据。数字图书馆从用户的注册信息中提取用户基础数据,借助网站自身的引导、调查及第三方提供等, 并在此基础上进行补充和交叉验证。自然维度的用户数据收集容易涉及用户隐私,因此自然维度的标签应是用户最基本的属性数据和影响用户使用数字图书馆的属性数据,笔者认为自然维度的标签体系应包括姓名、性别、年龄、学历、专业、职业、职称、城市,自然维度的标签相对静态,如果发生变化,允许用户自行修改。

3.2.2 兴趣维度　兴趣维度是用户画像的核心维度,旨在反映用户需求和兴趣,标签映射法的原理是通过为用户打标签,将用户兴趣特征化。兴趣维度的标签构建主要基于用户行为数据,用户在与数字图书馆的交互中产生了大量的用户服务日志,这些数据真实反映着用户的需求和兴趣偏好。兴趣维度的建立应首先对文本内容分析,构建资源画像,运用分词和去停用词得到特征词,并利用 TF-IDF 计算特征词权重,然后通过 LDA-Gibbs 模型对文本集建模,即利用文本的统计特性,将文本语料库映射到各个主题空间,挖掘隐藏在文本内的不同主题与词之间的关系,得到文档 - 主题概率分布和主题 - 特征词概率分布,构建资源画像标签,如图 1 所示:

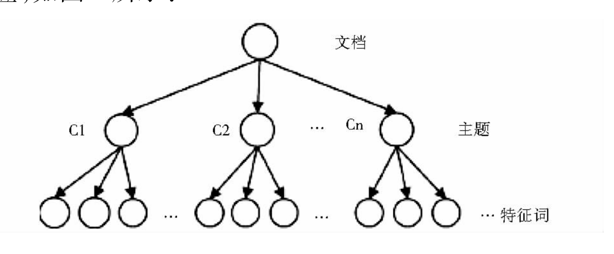


图 1　文档标签体系

然后进行用户兴趣识别。用户兴趣识别是指利用用户行为数据,建立用户 - 动词 - 对象三元组,计算不同交互类型对不同文档主题的兴趣强度,当用户与内容交互时,系统会聚合内容中标识的概念,并根据交互类型加权。具体方法如下:用 n 表示平台上用户的数

量,p 表示可能的交互类型的数量,m 表示平台上内容项的数量,k 表示内容中标识的主题的数量。用户兴趣主题矩阵如公式(1)所示:

$$UC_{n*k} = \sum_{v=1}^{p} w_v * UA_{n*m}^{v} * DC_{m*k} \quad \text{公式(1)}$$

其中,UC_{n*k} 是用户感兴趣主题的矩阵,因此 UC_{ij} 是用户 U_i 的主题 C_j 的相关性;w_v 是特定交互类型 v 的权重,表示用户的特定动作对内容的兴趣程度;UA_{n*m}^{v} 表示用户 - 内容的交互类型 v 矩阵;UA_{ij}^{v} 表示用户 U_i 完成 v 类型与内容 d_j 的交互的次数;DC_{m*k} 包含内容 - 主题分布,因此 DC_{f*r} 是主题 C_r 与内容项目 d_f 的相关性。

公式(1)适用于系统首次配置,但用户的交互是动态的,系统不断记录着用户行为数据,且用户的兴趣也在发生不断变化,所以在用户每次与内容交互中,要实时更新用户兴趣模型,如公式(2)所示:

$$UC_{1*k}^{after} = UC_{1*k}^{before} + w_v * UA_{n*m}^{v} * DC_{m*1} \quad \text{公式(2)}$$

其中,UC_{1*k}^{before} 表示本次交互前用户 - 兴趣主题矩阵,UC_{1*k}^{after} 表示本次交互后的用户 - 兴趣主题矩阵。依据用户兴趣主题矩阵,得到用户兴趣维度主题 - 特征词标签体系,根据其权重进行排序。

3.2.3 社交维度 社交维度在于构建用户关系网络图谱,数字图书馆本身具有虚拟性、交互性、开放性、自由性等特点,鼓励用户参与知识交流、共享、传播与创新,因此需要建立良好的知识交流社区。借鉴用户生成内容(UGC)社区,为用户与用户之间的知识获取和传递发挥了重要作用,用户在利用知识的同时也是传播和分享知识的载体,通过用户间的关注、咨询、讨论、分享、引用、被引、合作等互动行为,形成良好社交关系。分析用户群体属性,借助引证、合作等关系刻画个体用户间的社交、爱好、科研兴趣等的关联,揭示用户群间贡献度、活跃可见度等指数,形成不同类型、不同范围的动态关系网络图谱。将用户 ID 和用户互动数据导入复杂网络分析工具中,构建用户关系图谱,通过影响力分析筛选出核心用户,根据用户参与度和活跃度划分活跃用户、沉默用户、流失用户、回流用户,形成用户分级分类管理体系。另外,通过聚类算法发现相似用户,形成兴趣群体,进行用户推荐和专家推荐,建立兴趣小组,方便有共同兴趣用户之间的交流。

3.3 数字图书馆用户画像模型

由于认知方式的差异和用户画像构建宗旨的不同,国内外学者相继提出不同的用户画像构建方法。B. Rumpler 主张采用社会学方式、人机交互方式、认知方式、案例推理 4 种方式构建用户画像模型[39]。Y. Kritikou 提出用户画像模型建模的 3 个功能层:第一层是监控层,通过监控用户行为收集用户关键指标的值,形成用户资料;第二层是建模层,收集用户信息并以有效的方式对用户进行画像,建立用户间的关系;第三层是适应层,根据系统反馈,实施更新优化用户画像模型[40]。S. Henczel 将用户画像模型的构建分为 6 个阶段:①制定计划,确定需要收集的用户数据维度;②定位现有数据,识别所需的补充数据;③通过实地调研、访谈等方式开展调查收集资料;④分析用户行为数据,识别用户需求特征,并将具有共同需求特征的用户集群;⑤生成用户画像并对其进行评估测试;⑥在持续的交互和反馈中改进用户画像,完善相关服务[41]。

借鉴国内外数字图书馆用户画像模型,笔者认为数字图书馆用户画像本质是基于数据驱动将用户相关数据充分利用,将用户需求通过数据化→标签化→关联化→可视化展示出来,并应用于数字图书馆的各类型服务中,实现精准服务。数字图书馆画像模型构建流程大致可分为数据采集、数据预处理、数据存储、数据挖掘、形成用户画像、可视化及应用。数据采集模块主要是对用户注册信息的提取,通过问卷或访谈等完善用户基本属性数据,设置网页端和移动端的 API 接口记录用户交互数据以及挖掘 Web 日志,采集原始的用户相关数据。数据预处理模块是将采集的原始数据进行筛选和清洗,去除非相关数据,并运用数据处理工具将用户相关数据进行去重、去除非法字段、字段拆分、字段合并、资源数据信息代码表转换以及数据类型规范化等处理,将用户相关数据碎片化、数据化,形成标准的用户画像数据。数据存储模块是将处理好的用户画像数据集成和分类,形成用户基本属性数据库、用户行为数据库、用户互动数据库、用户其他数据库,并存储在 HBase 或 MongoDB 等分布式储存系统中,以便进行查询和计算。数据挖掘模块需要借助大数据工具 Hadoop 或 Spark 对数据库中的用户画像数据进行挖掘,计算用户兴趣权重,结合资源画像,建立用户标签体系,形成用户标签库,实现用户特征标签化。运用 Ucinet、Gephi 或 Pajek 等网络分析工具,建立用户关系图谱,并对用户进行分类和聚类,建立用户、专家、资源之间的联系,实现关联化。用户画像模块是将数据挖掘结果通过直方图、词云图、多维度多层级标签体系和关系图谱可视化展示出来,同时用户画像的结果接受用户的反馈,并根据用户的反馈做出调整和完善。应用模块是将用户画像结果应用在数字图书馆的知识服

务中，改善信息检索系统，优化检索结果排序，实现个性化检索；利用用户画像刻画的用户需求和兴趣，改进推荐系统，实现精准推荐；分析用户行为和群体特征，通过观测各群体活动进行针对性的宣传，为图书馆决策的实施提供参考。本文构建的数字图书馆用户画像框架模型如图 2 所示：

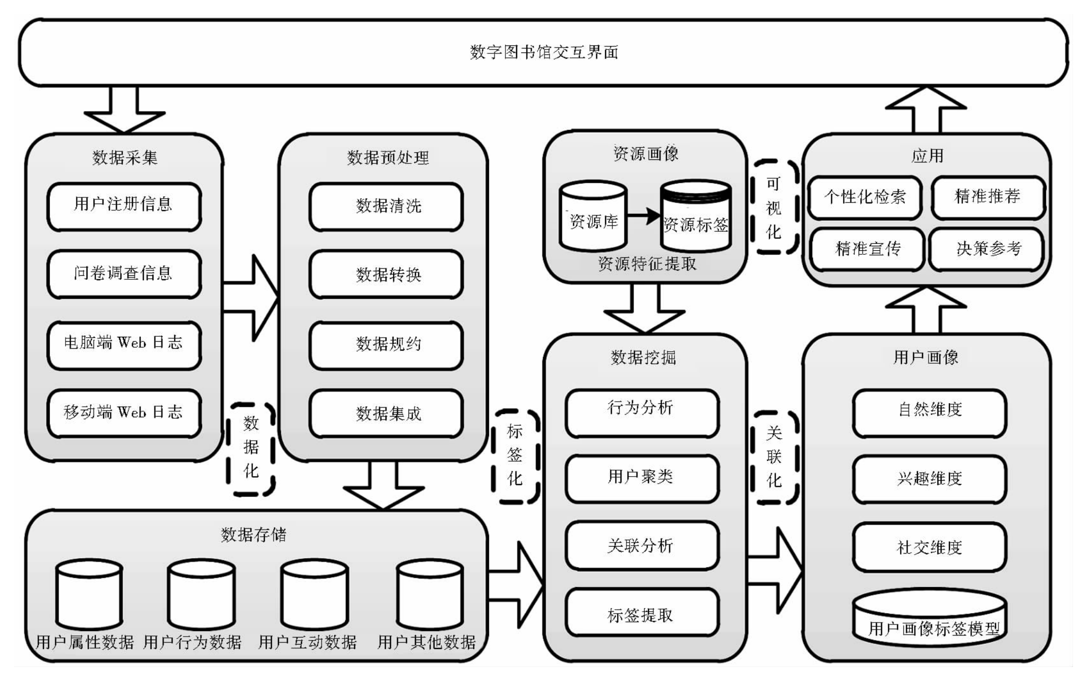


图 2　数字图书馆用户画像框架模型

4　数字图书馆用户画像的应用

4.1　精准推荐

传统的推荐算法，如基于内容的推荐算法从资源内容特征出发为用户进行推荐，而没有充分了解用户兴趣；基于协同过滤的推荐需要借助用户评分进行，而数字图书馆中用户主动性较差无法获得大量评分数据，且协同过滤算法具有稀疏性和冷启动的缺点。将用户画像应用于数字图书馆推荐系统中，形成对用户动态兴趣的识别和用户需求的预判，并结合资源画像，对数字馆藏资源进行碎片化挖掘、语义化描述、关联化连接，深度把握资源内容特征，从而实现用户画像准确的、动态的用户标签与数字图书馆资源深层语义标签的关联映射，结合相关性特征、情景特征、协同特征，全面升级推荐系统，并设计性能极致的召回策略，提高推荐结果的准确性，为用户实时准确地推荐符合其兴趣和需求的信息，智能地为用户提供个性化知识推送服务，提升知识发现服务能力，实现知识共享和知识价值的共创。

4.2　个性化检索

与传统文献检索方式不同，基于用户画像的个性化检索的特色功能在于通过分析用户检索历史的行为大数据，建立检索词、检索结果与用户行为数据的关联。构建数字图书馆智能检索系统，借助用户画像所提供的用户信息需求、检索行为、浏览习惯 、浏览主题等数据，通过大数据挖掘和分析为用户量身定制检索方式，为用户提供个性化检索服务。在检索结果的排序上，根据用户画像对用户兴趣和偏好的主题的把握和预测，将检索结果中用户最感兴趣、对用户最有价值的资源排在前列，既提高检索结果的准确率，又提高用户的满意度和交互黏性，更好地改善用户个性化检索体验，避免了无差别推送带来的资源冗余问题。

4.3　精准宣传

精准宣传不同于以往广撒网的宣传方式，而是以目标用户为中心，针对目标用户群体、受众群体进行宣传，更加注重宣传效果。用户画像基于大量的用户数据多维立体地展示了用户原貌，在此基础上借助分众分类、用户聚类、关联规则、数据挖掘和统计分析等工具，对用户进行全面的追踪和精细的划分，形成不同的用户群体，如根据用户使用习惯和活跃程度，将用户划分为活跃用户、沉默用户、流失用户、回流用户；根据用户参与程度及其影响力，分辨出核心用户、忠实用户、普通用户、潜在用户等[42]，准确地对用户进行分级管理；根据用户兴趣特征，可以将相似用户进行聚类，建立用户兴趣小组，推荐相似兴趣的用户和该兴趣领域的专家给用户，针对不同的用户群体采取不同的宣传

方式,有针对性地将宣传内容准确传递给受众群体,避免宣传浪费和对非相关用户造成干扰,提升宣传效果。在精准宣传策略实施过程中,数字图书馆可以通过与用户互动沟通、用户评价反馈等数据查看宣传服务的效果,并将这些效果反馈到用户画像数据库中,以便随时改进宣传策略。

4.4　参考决策

对用户而言,用户画像通过用户特征标签化的解读为用户自我认知提供参照,为用户的学习与研究、发展与进步提供决策依据。对数字图书馆而言,借助用户画像获取用户的特征信息和动态兴趣需求,以此来指导图书馆的服务导向和发展趋势,为数字图书馆采购资源、制定规章、提供服务等提供参考。通过用户画像分析用户对数字图书馆不同类型资源的需求,从而调整馆藏传统资源、半新资源、全新资源的配置比例。更为重要的是,通过可理解的用户标签和可识别的用户兴趣,用户的需求渐趋具象化。数字图书馆发展的根本目标是不断调整其服务内容和服务类型以提升用户满意度,用户画像的引入为数字图书馆提供了功能导向和服务依据。

5　结语

数据驱动环境下,数据作为一种资源,各行各业日益重视数据赋能,实现数据红利,用户数据是用户需求日趋碎片化、多元化、精细化的重要体现,如何挖掘用户数据,全面了解用户需求,发挥数据价值,是实现图书馆精准服务的关键点。基于用户数据构建全方位立体化的用户画像,为数字图书馆知识服务创新提供了新思路。对用户而言,其多元化、个性化需求得以满足,对数字图书馆而言,通过对用户特征的分析和动态兴趣的把握实现精准服务,提高用户满意度。本着双赢的目标,笔者对数字图书馆用户画像的内涵及特征进行了剖析,认为数据驱动下用户画像应围绕数据化→标签化→关联化→可视化为驱动的主线实施设计。依据驱动主线,笔者从自然维度、兴趣维度、社交维度构建了多维度多层级立体化的用户画像模型,设计了数字图书馆用户画像的模型框架,阐述了用户画像在数字图书馆精准推荐、个性化检索、精准宣传、参考决策等方面的具体应用,有助于打破大数据蓝海中信息孤岛的桎梏,促进数字图书馆知识服务的创新优化,为数字图书馆转型升级、提升知识发现服务能力、实现精准服务提供新动能。

参考文献:

[1] 毕强,闫晶,李洁. 大数据时代数字图书馆服务转型面临的新形势与新要求[J]. 情报理论与实践,2017,40(12):12-16,5.
[2] 库珀. 交互设计之路[M]. 北京:电子工业出版社,2006:10.
[3] 亓丛,吴俊. 用户画像概念溯源与应用场景研究[J]. 重庆交通大学学报(社会科学版),2017,17(5):82-87.
[4] MARSHALL R,COOK S,MITCHELL V,et al. Design and evaluation: end users, user datasets and personas. [J]. Applied ergonomics,2015,46(B):311-317.
[5] TEIXEIRA C,PINTO J S,MARTINS J A. User profiles in organizational environments. [J]. Campus-wide information systems,2015,25(25):329-332.
[6] ZENG L, ZHANG Y, QIU R G. Adaptive user profiling in enhancing RSS-based information services[C]//IEEE international conference on service operations and logistics, and informatics. Philadelphia:IEEE,2007:1-5.
[7] AMO S D, DIALLO M S, DIOP C T, et al. Contextual preference mining for user profile construction[J]. Information systems,2015,49(C):182-199.
[8] ACHEMOUKH F, AHMED-OUAMER R. Representation and evolution of user profile in information retrieval based on Bayesian approach[C]// International symposium on methodologies for intelligent systems. Cham:Springer international publishing,2014:486-492.
[9] YU J, LIU F, ZHAO H. Building User profile based on concept and relation for web personalized services[C]//International conference on innovation and information management. Singapore:ICI-IM,2012(36):165-172.
[10] 陈添源. 高校移动图书馆用户画像构建实证[J]. 图书情报工作,2018,62(7):38-46.
[11] AL-SHAMRI M Y H. User profiling approaches for demographic recommender systems[J]. Knowledge-based systems,2016(100):175-187.
[12] 付小飞. 基于用户画像的移动广告推荐技术的研究与应用[D]. 成都:电子科技大学,2017.
[13] 单晓红,张晓月,刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践,2018,41(4):99-104,149.
[14] 赵曙光. 高转化率的社交媒体用户画像:基于500用户的深访研究[J]. 现代传播(中国传媒大学学报),2014,36(6):115-120.
[15] SHIRUDE S B, KOLHE S R. Measuring similarity between user profile and library book[C]// International conference on information systems and computer networks. Mathura:IEEE,2014:50-54.
[16] 马费成,周利琴. 面向智慧健康的知识管理与服务[J]. 中国图书馆学报,2018(5):1-15.
[17] BISHOP J, LEWIS P R. BLAISE-LINE and the British National Bibliography:profiles of users and uses[J]. Journal of librarianship & information science,1985(2):119-136.
[18] AMATO G, STRACCIA U. User profile modeling and applications to digital libraries[C]// European conference on research and advanced technology for digital libraries. Berlin:Springer,1999:184-197.
[19] ZHE S L,DONG N Z,TAO S H,et al. Study on the user profile of personalized service in digital Library[J]. Journal of Beijing Insti-

tute of Technology,2005,25(1):58－62.

[20] MAO J, LU K, LI G, et al. Profiling users with tag networks in diffusion-based personalized recommendation[J]. Journal of information science,2016,42(5):711－722.

[21] SHARMA D, KAUR S. Neural network classification for user profile learning over digital library recommendation engine[J]. Indian journal of science & technology,2016,9(33):1－7.

[22] ZAUGG H. Using persona descriptions to inform library space design[M]//The future of library space. Bradford: Emerald Group Publishing Limited,2016:335－358.

[23] 刘速.浅议数字图书馆知识发现系统中的用户画像——以天津图书馆为例[J].图书馆理论与实践,2017(6):103－106.

[24] 杨帆.以画像分析为基础的图书馆大数据——以国家图书馆大数据项目为例[J].图书馆论坛,2019(2):1－8.

[25] 何娟.基于用户个人及群体画像相结合的图书个性化推荐应用研究[OL].[2018－08－20].http://kns.cnki.net/kcms/detail/11.1762.G3.20180816.1745.009.html.

[26] 尹相权,李书宁,弓建华.基于系统日志的高校图书馆研究间用户利用行为分析[J].现代情报,2018,38(1):115－120.

[27] 薛欢雪.高校图书馆学科服务用户画像创建过程[J].图书馆学研究,2018(13):67－71,82.

[28] 胡媛,毛宁.基于用户画像的数字图书馆知识社区用户模型构建[J].图书馆理论与实践,2017(4):82－85,97.

[29] 王凌霄,沈卓,李艳.社会化问答社区用户画像构建[J].情报理论与实践,2018,41(1):129－134.

[30] 范晓玉, 窦永香, 赵捧未,等. 融合多源数据的科研人员画像构建方法研究[J]. 图书情报工作, 2018, 62(15): 31－40.

[31] 陈慧香,邵波.国外图书馆领域用户画像的研究现状及启示[J].图书馆学研究,2017(20):16－20.

[32] TRAVIS D. E－Commerce usability: tools and techniques to perfect the onLine experience[M]. Oxford:Routledge,2002.

[33] FARSEEV A, NIE L, AKBARI M, et al. Harvesting multiple sources for user profile learning: a big data study[C]// Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Shanghai:ACM,2015:235－242.

[34] GUIMARAES T P. Perfil de usuários de biblioteca governamental: O Caso do ministério da saúde[J]. Perspectivas em ciência da informa92o,2007(3):96－115.

[35] LAFOUGE T, LARDY J P, ABDALLAH N B. Improving information retrieval by combining user profile and document segmentation[J]. Information processing & management,1996,32(3):305－315.

[36] 汪强兵,章成志.融合内容与用户手势行为的用户画像构建系统设计与实现[J].数据分析与知识发现,2017,1(2):80－86.

[37] 张慧敏,辛向阳.构建动态用户画像的四个维度[J].工业设计,2018(4):59－61.

[38] 陈志明,胡震云.网站用户画像研究[J].计算机系统应用,2017,26(1):24－30.

[39] RUMPLER B. A study of the impact of the user profile in documentary systems[J]. Online information review,2001,25(6):359－365.

[40] KRITIKOU Y, DEMESTICHAS P, ADAMOPOULOU E, et al. User profile modeling in the context of web-based learning management systems[J]. Journal of network & computer applications, 2008,31(4):603－627.

[41] HENCZEL S. Creating user profiles to improve information quality[J]. Online, 2004, 28(3):30－33.

[42] 曾建勋.精准服务需要用户画像[J].数字图书馆论坛,2017(12):1.

作者贡献说明：

许鹏程：提出研究命题、撰写论文；

毕强：提出研究思路，修改论文；

张晗：英文摘要撰写，修改论文；

牟冬梅：完善论文思路，修改论文。

Construction of Digital Library User Profile Driven by Data

Xu Pengcheng[1] Bi Qiang[1] Zhang Han[1] Mu Dongmei[2]

[1] School of Management, Jilin University, Changchun 130022

[2] School of Public Health, Jilin University, Changchun 130021

Abstract: [**Purpose/significance**] This paper designs a digital library user profile model, in order to explore the hidden value behind user data, comprehensively understand the needs of users, and provide new kinetic energy for the digital library to achieve precise services. [**Method/process**] This paper analyzes the connotation and characteristics of the user profile of the digital library, analyzes the data source and collection process of the user's profile, and regards its driven main route as digitization, labeling, association and visualization. From the natural dimension, interest dimension and social dimension, the article constructs a multi-dimensional user profile model. [**Result/conclusion**] The paper describes the construction process of the user profile model and designs a model framework for user profile. Simultaneously this article applies the user profile to the precise recommendation, personalized retrieval, accurate publicity and reference decision-making to promote the digital library's knowledge service upgrade.

Keywords: data driven digital library user profile